

# A new upper bound on the work function algorithm for the *k*-server problem

Wenming Zhang<sup>1</sup> · Yongxi Cheng<sup>2,3</sup>

Published online: 28 November 2019 © Springer Science+Business Media, LLC, part of Springer Nature 2019

# Abstract

The *k*-server problem was introduced by Manasse et al. (in: Proceedings of the 20th annual ACM symposium on theory of computing, Chicago, Illinois, USA, pp 322–333, 1988), and is one of the most famous and well-studied online problems. Koutsoupias and Papadimitriou (J ACM 42(5):971–983, 1995) showed that the work function algorithm (*WFA*) has a competitive ratio of at most 2k - 1 for the *k*-server problem. In this paper, by proposing a potential function that is different from the one in Koutsoupias and Papadimitriou (1995), we show that the *WFA* has a competitive ratio of at most n - 1, where *n* is the number of points in the metric space. When n < 2k, this ratio is less than 2k - 1.

**Keywords** Competitive analysis  $\cdot k$ -server problem  $\cdot$  On-line algorithms  $\cdot$  Work function

# **1 Introduction**

The *k*-server problem was introduced by Manasse et al. (1988), and is one of the most famous and well-studied online problems (for an excellent reference, please refer to Koutsoupias 2009). The problem is defined on a metric space  $\mathcal{U}$ , which is a set of *n* points with a symmetric distance function *d* (nonnegative real function) that satisfies the triangle inequality. On the metric space  $\mathcal{U}$ , *k* servers reside that can move from point to point.

Yongxi Cheng chengyx@mail.xjtu.edu.cn
 Wenming Zhang wenming@nwu.edu.cn

<sup>&</sup>lt;sup>1</sup> School of Economics and Management, Northwest University, Xi'an 710127, China

<sup>&</sup>lt;sup>2</sup> School of Management, Xi'an Jiaotong University, Xi'an 710049, China

<sup>&</sup>lt;sup>3</sup> State Key Lab for Manufacturing Systems Engineering, Xi'an 710049, China

As in Koutsoupias and Papadimitriou (1995), a possible position of the *k* servers is called a *configuration*, which can be viewed as a multiset of *k* points of  $\mathcal{U}$ . We use capital letters for configurations. For two configurations *X* and *Y*, we use D(X, Y) for the minimum distance to move the servers from *X* to *Y*. We assume that the *k* servers are initially at a fixed configuration  $A_0$ .

A request sequence  $\rho$  is a sequence of points of the metric space  $\mathcal{U}$  to be serviced by the *k* servers; servicing a request entails moving some server to the point of request. Let  $opt(A_0, \rho)$  denote the optimal off-line cost for servicing a request sequence  $\rho$  starting at the initial configuration  $A_0$ , and let  $cost(A_0, \rho)$  denote the cost for servicing  $\rho$  of some online algorithm. The competitive ratio of the online algorithm is the infimum of all *c* such that for all initial configurations  $A_0$  and for all request sequences  $\rho$ 

$$cost(A_0, \rho) \leq c \cdot opt(A_0, \rho) + C$$

where *C* may depend on the initial configuration  $A_0$  but not on the request sequence  $\rho$ . An online algorithm with competitive ratio *c* is called *c*-*competitive*. For an overview on competitive analysis and online algorithms, please refer to Komm (2016).

It has been shown that for the *k*-server problem no deterministic online algorithm has a competitive ratio less than *k* (Manasse et al. 1988). The famous *k*-server conjecture, which has been open for over 30 years, states that there exists a *k*-competitive online algorithm for every metric space (Manasse et al. 1988). Currently, the top candidate online algorithm for settling this conjecture is the *work function algorithm (WFA)*, which was shown to have a competitive ratio of at most 2k - 1 (Koutsoupias and Papadimitriou 1995).

The WFA was first made explicit in the work of Chrobak and Larmore (1992), where the algorithm was proved to be 2-competitive for k = 2. Then, Koutsoupias and Papadimitriou (1995) showed that the WFA has a competitive ratio of at most 2k - 1, for any  $k \ge 1$ . They conjecture that the WFA is in fact k-competitive. To support this conjecture, Bartal and Koutsoupias (2004) proved that the WFA achieves ratio k in several special metric spaces: the line, the star, and all metric spaces with k + 2 points. In addition, numerical results from several implementations of the WFA (Rudec et al. 2010, 2013; Rudec and Manger 2015) indicate that, the WFA could indeed assure better cost than simple heuristics, such as the greedy algorithm or the balanced algorithm.

In this paper, by proposing a potential function that is different from the one in Koutsoupias and Papadimitriou (1995), we show that the *WFA* has a competitive ratio of at most n - 1, where n is the number of points in metric space U. When n < 2k, this ratio is less than 2k - 1.

#### 2 Preliminaries

Our work is mainly based on the idea and tools developed in Koutsoupias and Papadimitriou (1995), and we adopt the notation and definitions therein. To make the paper more self-contained, we list them as the following.

*Work Function* Fix a metric space  $\mathcal{U}$  and an initial configuration  $A_0$ . For a request sequence  $\rho$  define the work function  $w_\rho$  from configurations to nonnegative real num-

bers as follows:  $w_{\rho}(X)$  is the optimal cost of servicing  $\rho$  starting at  $A_0$  and ending up at configuration X.

The subscript  $\rho$  is usually omitted from  $w_{\rho}$  when it is clear from the context. Furthermore, for a work function  $w = w_{\rho}$ ,  $w' = w_{\rho r}$  is referred to as the resulting work function after request r, when  $\rho$  and r are understood from the context. The value of w'(X) can be computed as follows (which is stated as Fact 1 in Koutsoupias and Papadimitriou 1995)

$$w'(X) = \min_{x \in X} \{ w(X - x + r) + d(r, x) \}.$$

For convenience, in this paper we also refer to the above property of the work function as Fact 1. Also, by Fact 4 in Koutsoupias and Papadimitriou (1995), for all configurations  $X, w'(X) \ge w(X)$ .

Work Function Algorithm (WFA) Let  $\rho$  be a request sequence and let A be the configuration of an online algorithm after servicing  $\rho$ . The work function algorithm services a new request r by moving its servers to a configuration A', with  $r \in A'$ , that minimizes  $w_{\rho r}(A') + D(A, A')$ .

*Extended Cost* The extended cost for request r is defined to be the maximum increase of the work function after request  $r: \max_X \{w'(X) - w(X)\}$ . The extended cost is said to *occur* on a configuration A when A maximizes the quantity in the extended cost.

It is showed that if the total extended cost of all requests is bounded above by c + 1 times the optimal off-line cost plus a constant, then the *WFA* is *c*-competitive (Fact 5 in Koutsoupias and Papadimitriou 1995).

*Minimizer* A configuration A is called a minimizer of a point a with respect to w, if A minimizes the expression  $w(X) - \sum_{x \in X} d(a, x)$ , that is

$$w(A) - \sum_{x \in A} d(a, x) = \min_{X} \{ w(X) - \sum_{x \in X} d(a, x) \}.$$

The following lemma, which is proved in Koutsoupias and Papadimitriou (1995), is crucial in later analysis. For convenience of the readers, we put the proof of this lemma in "Appendix A", to make the paper more self-contained.

**Lemma 2.1** (Duality lemma) Let w be a work function and w' be the resulting work function after request r. Then any minimizer A of r with respect to w is also a minimizer of r with respect to w', and the extended cost of servicing the request r occurs on A. That is,

$$w'(A) - w(A) = \max_{X} \{ w'(X) - w(X) \}.$$

#### 3 A potential for (n - 1)-competitiveness

In this section, we will show that the WFA is (n-1)-competitive by proposing a potential different from that in Koutsoupias and Papadimitriou (1995). Let  $u_1, u_2, \ldots, u_n$  denote the *n* points in metric space  $\mathcal{U}$ . For configurations  $B_i = \{b_{i1}, b_{i2}, \dots, b_{ik}\}, i = 1, 2, \dots, n$ , let

$$\Psi(w, B_1, B_2, \dots, B_n) = \sum_{i=1}^n \left( w(B_i) - \sum_{j=1}^k d(u_i, b_{ij}) \right).$$

Let  $\Phi(w)$  denote its minimum value over all configurations  $B_i$ , i = 1, 2, ..., n;  $\Phi(w)$  is called the *potential* of the work function w.

**Lemma 3.1** Let  $A_0 = \{a_1^0, a_2^0, \dots, a_k^0\}$ . For the initial work function  $w_e(X) = D(A_0, X)$ ,

$$\Phi(w_e) = -\sum_{i=1}^{n} \sum_{j=1}^{k} d(u_i, a_j^0).$$

**Proof** It is easy to see that the lemma follows if the minimum value  $\Phi(w_e)$  of  $\Psi(w_e, B_1, B_2, \dots, B_n)$  is achieved when  $B_i = A_0$  for  $i = 1, 2, \dots, n$ .

Consider a point  $b_{ij} \in B_i$ . In the minimum matching  $D(A_0, B_i)$ ,  $b_{ij}$  is matched to some point  $a \in A_0$ . By using the triangle inequality  $d(u_j, b_{ij}) \leq d(a, u_j) + d(a, b_{ij})$ , we see that we can replace  $b_{ij}$  with a without increasing the value of  $\Psi(w_e, B_1, B_2, \ldots, B_n)$ . Therefore the minimum value  $\Phi(w_e)$  of  $\Psi(w_e, B_1, B_2, \ldots, B_n)$  is achieved when  $B_i = A_0$ , for  $i = 1, 2, \ldots, n$ . The lemma follows.

Now we are ready to prove our main result.

**Theorem 3.1** The competitive ratio of the WFA is at most n - 1, where n is the number of points in metric space U.

**Proof** Consider a work function w and let w' be the resulting work function after request r. We assume that  $\Psi(w', B_1, B_2, \ldots, B_n)$  achieves its minimum with  $B_1^*, B_2^*, \ldots, B_n^*$ , that is,

$$\Phi(w') = \Psi(w', B_1^*, B_2^*, \dots, B_n^*).$$

Let  $B_i^* = \{b_{i1}^*, b_{i2}^*, \dots, b_{ik}^*\}$ , for  $i = 1, 2, \dots, n$ .

Assume that  $r = u_l$  for some  $l \in \{1, 2, ..., n\}$ . Let A be a minimizer of r with respect to w. Then, by Lemma 2.1, A is also a minimizer of r with respect to w', and it is not difficult to see that the minimum value of  $\Psi(w', B_1, B_2, ..., B_n)$  is unaffected if we fix  $B_l = A$ . Hence, without loss of generality, we assume that  $B_l^* = A$ . Furthermore, from the definition of  $\Phi(w)$ , we get that

$$\Phi(w) \leq \Psi(w, B_1^*, B_2^*, \dots, B_n^*).$$

Thus,

$$\Phi(w') - \Phi(w) \ge \Psi(w', B_1^*, B_2^*, \dots, B_n^*) - \Psi(w, B_1^*, B_2^*, \dots, B_n^*)$$

$$=\sum_{i=1}^{n} \left( w'(B_i^*) - w(B_i^*) \right)$$
$$\geq w'(A) - w(A),$$

where the last inequality holds since  $w'(B_i^*) \ge w(B_i^*)$  for i = 1, 2, ..., n and  $B_i^* = A$ .

According to Lemma 2.1, the extended cost is w'(A) - w(A), because A is a minimizer of r with respect to w. Thus, we can conclude that the extended cost to service request r is bounded above by  $\Phi(w') - \Phi(w)$ . Summing over all moves, the total extended cost is bounded above by  $\Phi(w_{\rho}) - \Phi(w_e)$ , where  $w_e$  and  $w_{\rho}$  are the initial and the final work functions, respectively.

Let  $A_0 = \{a_1^0, a_2^0, \dots, a_k^0\}$  and  $A_t = \{a_1^t, a_2^t, \dots, a_k^t\}$  be the initial and final configurations (as in Koutsoupias and Papadimitriou (1995), without loss of generality we can assume that the off-line algorithm ends up in the same configuration as the online algorithm). From the definition of  $\Phi(w_\rho)$ , we have

$$\Phi(w_{\rho}) \leq \Psi(w_{\rho}, A_t, A_t, \dots, A_t)$$
$$= \sum_{i=1}^n \left( w_{\rho}(A_t) - \sum_{j=1}^k d(u_i, a_j^t) \right)$$
$$\leq n w_{\rho}(A_t).$$

The value of  $\Phi(w_e)$  is given by Lemma 3.1, which is a constant (depending only on the initial configuration). Hence, the total extended cost is bounded above by  $nw_\rho(A_t)$  plus a constant. Because the optimal off-line cost is  $w_\rho(A_t)$ , the total extended cost is bounded above by n times the off-line cost plus a constant, implying that the *WFA* is (n-1)-competitive.

By combining the result that the WFA is (2k - 1)-competitive in Koutsoupias and Papadimitriou (1995), we have the following result.

**Corollary 3.1** *The competitive ratio of the* WFA *is at most*  $min\{2k - 1, n - 1\}$ *.* 

#### 4 Conclusion

In this paper, we propose a potential function that is different from the one in Koutsoupias and Papadimitriou (1995), and by which we show that the *WFA* for the *k*-server problem has a competitive ratio of at most n - 1, where *n* is the number of points in metric space  $\mathcal{U}$ . When n < 2k, this ratio is less than 2k - 1. Whether or not the *WFA* has a competitive ratio of *k* for the *k*-server problem remains a central open problem in the field of on-line algorithms.

Acknowledgements This work was partially supported by Special Scientific Research Plan of Shaanxi Province Education Department (17JK0785) and the National Natural Science Foundation of China under Grant No. 11771346. The authors would like to thank the reviewer for helpful comments and suggestions.

### Appendix A: Proof of Lemma 2.1

The following is the proof of Lemma 2.1 given in Koutsoupias and Papadimitriou (1995).

**Definition A.1** A function w is called quasiconvex if for all configurations A, B, there exists a bijection  $h : A \rightarrow B$  such that for all bipartitions of A into X, Y:

$$w(A) + w(B) \ge w(X \cup h(Y)) + w(h(X) \cup Y).$$

$$(1)$$

Notice that the union  $(\cup)$  in the definition denotes the union of multisets. Before showing that all work functions are quasiconvex, the following lemma provides a stronger form of the quasiconvexity condition by restricting the set of possible bijections.

**Lemma A.1** If there exists a bijection h that satisfies the conditions in Definition A.1, then there exists a bijection h' that satisfies the same conditions and h'(x) = x for all  $x \in A \cap B$ .

**Proof** Let *h* be a bijection from *A* to *B* that satisfies the conditions of the definition above and maps the maximum number of elements in  $A \cap B$  to themselves. Assume that, for some  $a \in A \cap B$ , we have  $h(a) \neq a$ . Define a bijection h' that agrees with *h* everywhere except that

$$h'(a) = a$$
 and  $h'(h^{-1}(a)) = h(a)$ 

(h' interchanges the values of h on a and  $h^{-1}(a)$ ).

Consider now a bipartition of *A* into *X* and *Y* and assume (without loss of generality) that  $h^{-1}(a) \in X$ . If  $a \in X$ , then h(X) = h'(X) and h(Y) = h'(Y) and (1) holds for *h'*. Otherwise, when  $a \notin X$ , we derive the quasiconvexity condition for *X* and *Y* from the quasiconvexity condition for X' = X + a and Y' = Y - a as follows:

Since, h(Y') = h'(Y') and h(X') = h'(X'), we have that

$$X' \cup h(Y') = X' \cup h'(Y') = (X+a) \cup h'(Y-a) = X \cup h'(Y),$$

and similarly,  $h(X') \cup Y' = h'(X) \cup Y$ . From these, we get

$$w(A) + w(B) \ge w(X' \cup h(Y')) + w(h(X') \cup Y')$$
  
=  $w(X \cup h'(Y)) + w(h'(X) \cup Y).$ 

Therefore, h' satisfies the quasiconvexity condition. Because h' maps at least one more element in  $A \cap B$  to itself than h, it contradicts the assumption that h maps the maximum number of elements in  $A \cap B$  to themselves.

We conclude that h(a) = a for all  $a \in A \cap B$ , and the lemma holds.

Lemma A.2 (Quasiconvexity lemma) All work functions are quasiconvex.

**Proof** The proof is by induction on the number of requests.

Recall that the initial work function  $w_e(X)$  of a configuration X is equal to  $D(A_0, X)$ , where  $A_0$  is the initial configuration. So we have

$$w(A) + w(B) = D(A_0, A) + D(A_0, B).$$

Fix two matchings  $M(A_0, A)$  and  $M(A_0, B)$  that realize the minima of  $D(A_0, A)$  and  $D(A_0, B)$ . Each point  $x_j$  in  $A_0$  is matched to some point  $a_j$  in A and  $b_j$  in B. Consider the bijection  $h : A \to B$  that maps each  $a_j$  to  $b_j$ . For any bipartition of A into X and Y, w(X + h(Y)) + w((h(X) + Y)) is equal to the sum of two minima matchings between  $A_0, X + h(Y)$  and  $A_0, h(X) + Y$ . Since we can rearrange the matchings  $M(A_0, A)$  and  $M(A_0, B)$  to obtain two matchings (not necessarily minima) between  $A_0, X + h(Y)$  and  $A_0, h(X) + Y$ , it follows that  $w(A) + w(B) \ge w(X + h(Y)) + w(h(X) + Y)$ .

For the induction step, assume that w is quasiconvex. We want to show that the resulting w' after request r is also quasiconvex.

Fix two configurations A and B. Using Fact 1 to express w' in terms of w, we get that w'(A) = w(A - a + r) + d(r, a) for some  $a \in A$ ; similarly w'(B) = w(B - b + r) + d(r, b) for some  $b \in B$ . The induction hypothesis is that w is quasiconvex, so there exists a bijection h from A - a + r to B - b + r that satisfies the quasiconvexity condition. Furthermore, Lemma A.1 allows us to assume that h(r) = r.

Consider now a bijection  $h' : A \to B$  that agrees with h everywhere, except that h'(a) = b. We show that h' satisfies the requirements of the quasiconvexity condition of w'. Consider a bipartition of A into X and Y and without loss of generality assume that  $a \in X$ . We have:

$$\begin{split} w'(A) + w'(B) &= w(A - a + r) + w(B - b + r) + d(r, a) + d(r, b) \\ &= w((X - a + r) \cup Y) + w(B - b + r) + d(r, a) + d(r, b) \\ &\geq w((X - a + r) \cup h(Y)) + w(h(X - a + r) \cup Y) \\ &+ d(r, a) + d(r, b) \\ &= w((X - a + r) \cup h'(Y)) + w((h'(X) - b + r) \cup Y) \\ &+ d(r, a) + d(r, b) \\ &\geq w'(X \cup h'(Y)) + w'(h'(X) \cup Y), \end{split}$$

where the first inequality is based on the quasiconvexity of w and the second one on Fact 1. So, w' is quasiconvex and the lemma follows.

Now we use the quasiconvexity condition to prove the next two lemmata. In fact, the following weaker condition derived from quasiconvexity is used:

$$\forall a \in A \exists b \in B : w(A) + w(B) \ge w(A - a + b) + w(B - b + a).$$

**Lemma A.3** Let w be a work function. Consider a new request at r and the resulting work function w'. If A is a minimizer of r with respect to w, then A is also a minimizer of r with respect to w'.

**Proof** It suffices to show that for all configurations B:

$$w'(B) - \sum_{b \in B} d(r, b) \ge w'(A) - \sum_{a \in A} d(r, a),$$

or equivalently:

$$w'(B) - \sum_{b \in B} d(r, b) + w(A) \ge w'(A) - \sum_{a \in A} d(r, a) + w(A).$$

In order to show this we need the following: From Fact 1, we get that there exists  $b' \in B$  such that

$$w'(B) = w(B - b' + r) + d(r, b').$$

Using quasiconvexity, we get that there exists  $a' \in A$  such that

$$w(B - b' + r) + w(A) \ge w(B - b' + a') + w(A - a' + r).$$

Finally, since A is a minimizer of r, we have that

$$w(B - b' + a') - \sum_{b \in B - b' + a'} d(r, b) \ge w(A) - \sum_{a \in A} d(r, a).$$

Putting all these together:

$$\begin{split} &w'(B) + w(A) - \sum_{b \in B} d(r, b) \\ &= w(B - b' + r) + d(r, b') + w(A) - \sum_{b \in B} d(r, b) \\ &= w(B - b' + r) + w(A) - \sum_{b \in B - b' + r} d(r, b) \\ &\geq w(B - b' + a') + w(A - a' + r) - \sum_{b \in B - b' + r} d(r, b) \\ &= w(B - b' + a') + w(A - a' + r) + d(r, a') - \sum_{b \in B - b' + a'} d(r, b) \\ &\geq w(A) + w(A - a' + r) + d(r, a') - \sum_{a \in A} d(r, a) \\ &\geq w(A) + w'(A) - \sum_{a \in A} d(r, a), \end{split}$$

where the last inequality is based on Fact 1. The lemma follows.

The following lemma has the same premises with Lemma A.3, but a different conclusion:

**Lemma A.4** Let w be a work function. Consider a new request at r and the resulting work function w'. If A is a minimizer of r with respect to w, then the extended cost occurs at A, that is

$$w'(A) - w(A) = \max_{X} \{ w'(X) - w(X) \}.$$

**Proof** The proof is rather similar to the proof of Lemma A.3. Notice first that it suffices to show that for all configurations *B*:

$$w'(A) + w(B) \ge w'(B) + w(A).$$

By Fact 1, we get that there exists  $a' \in A$  such that

$$w'(A) = w(A - a' + r) + d(r, a').$$

Using quasiconvexity, we also get that there exists  $b' \in B$  such that

$$w(A - a' + r) + w(B) \ge w(A - a' + b') + w(B - b' + r).$$

Finally, since A is a minimizer of r with respect to w:

$$w(A - a' + b') - \sum_{a \in A - a' + b'} d(r, a) \ge w(A) - \sum_{a \in A} d(r, a),$$

which is equivalent to

$$w(A - a' + b') + d(r, a') \ge w(A) + d(r, b').$$

Combining all these we get:

$$w'(A) + w(B) = w(A - a' + r) + d(r, a') + w(B)$$
  

$$\geq w(A - a' + b') + d(r, a') + w(B - b' + r)$$
  

$$\geq w(A) + d(r, b') + w(B - b' + r)$$
  

$$\geq w(A) + w'(B).$$

Again, the last inequality is based on Fact 1.

Lemmata A.3 and A.4 can be combined into Lemma 2.1, which characterizes where the extended cost occurs. The proof of Lemma 2.1 is established.

# References

Bartal Y, Koutsoupias E (2004) On the competitive ratio of the work function algorithm for the *k*-server problem. Theor Comput Sci 324(2–3):337–345

Chrobak M, Larmore LL (1992) The server problem and on-line games. In: On-line algorithms, DIMACS series in discrete mathematics and theoretical computer science, vol 7

Komm D (2016) An introduction to online computation. Springer International Publishing, Cham Koutsoupias E (2009) The *k*-server problem. Comput Sci Rev 3(2):105–118

Koutsouplas E (2007) The k-server problem. Comput Ser Kev 5(2).105-118

Koutsoupias E, Papadimitriou CH (1995) On the k-server conjecture. J ACM 42(5):971-983

- Manasse MS, Mcgeoch LA, Sleator DD (1988) Competitive algorithms for on-line problems. In: Proceedings of the 20th annual ACM symposium on theory of computing, 2–4 May 1988, Chicago, Illinois, USA, pp 322–333
- Rudec T, Manger R (2015) A fast approximate implementation of the work function algorithm for solving the *k*-server problem. Cent Eur J Oper Res 23(3):699–722
- Rudec T, Baumgartner A, Manger R (2010) Measuring true performance of the work function algorithm for solving the on-line *k*-server problem. J Comput Inf Technol 18(4):361–367
- Rudec T, Baumgartner A, Manger R (2013) A fast work function algorithm for solving the *k*-server problem. Cent Eur J Oper Res 21(1):187–205

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.